

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Il metodo sperimentale

This is a pre print version of the following article:

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1621631> since 2017-01-12T18:30:42Z

Publisher:

De Agostini UTET

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

Il metodo sperimentale

Davide Barrera

CAPITOLO 5

Uno dei primi problemi che vengono trattati nei corsi introduttivi di metodologia quantitativa è quello della correlazione spuria. La correlazione spuria si verifica tutte le volte che due fenomeni appaiono essere correlati, senza che vi sia una relazione di causa ed effetto fra i due. Il problema delle relazioni spurie è molto comune nelle ricerche basate su inchieste campionarie, perché nella maggior parte dei casi il disegno di queste ricerche è *cross-section*, cioè condotto in un determinato tempo, su una porzione di popolazione¹. In tutti i manuali o saggi di natura metodologica in cui si descrive il metodo sperimentale, quest'ultimo viene indicato come il metodo di elezione per determinare empiricamente relazioni di tipo causale (si vedano per esempio Willer e Walker 2007, Meeker e Leik 2007). L'esperimento è quella tecnica di ricerca per mezzo della quale il ricercatore controlla il fenomeno oggetto di studio stabilendo le condizioni in base alle quali quel dato fenomeno viene osservato e misurato. Il controllo di norma si manifesta mediante la creazione di due o più condizioni, virtualmente identiche tranne che per un singolo aspetto, che viene introdotto o manipolato dal ricercatore in una delle condizioni per osservarne gli effetti mediante confronto con la condizione in cui la manipolazione è assente, in modo tale da permettere una chiara inferenza causale. Convenzionalmente il termine esperimento viene usato anche per studi con caratteristiche diverse, per esempio i cosiddetti esperimenti di *rottura* (Garfinkel 1967). In questo capitolo verrà descritto il metodo sperimentale con riferimento ad esperimenti in cui sono presenti due o più condizioni create o controllate dal ricercatore; questa definizione accomuna gli esperimenti di laboratorio condotti in psicologia sociale, sociologia, ed economia sperimentale ed anche i cosiddetti esperimenti di campo (*field experiments*), laddove la manipolazione sperimentale che genera e distingue le due o più condizioni sia analoga a quelle che vengono implementate in laboratorio.

¹ Dal momento che la causa precede sempre l'effetto, l'inferenza di una relazione causale mediante inchiesta è meno problematica se il disegno della ricerca è longitudinale, cioè se la variabile causa viene rilevata in un tempo antecedente a quello in cui viene rilevata la variabile effetto.

5.1 Dai canoni di Mills all'inferenza causale

La relazione fra la tecnica dell'esperimento di laboratorio e il problema della determinazione della causalità viene solitamente fatta risalire ai canoni di John Stuart Mill ([1843] 1973). Mill elencò cinque metodi, noti come «i canoni di Mill», per stabilire la presenza di una relazione causale fra due fenomeni. Il primo è il *metodo dell'accordo*, in base al quale se il fenomeno oggetto di studio si verifica in due circostanze che hanno un solo elemento in comune, si può inferire che l'elemento comune ai due eventi ne costituisca la causa. Il secondo è il *metodo della differenza*, in base al quale se, date due situazioni identiche sotto ogni aspetto con l'eccezione di un solo elemento, il fenomeno oggetto di studio si verifica solo nella situazione in cui l'elemento distintivo è presente, si può inferire che quell'elemento costituisca la causa del fenomeno. Il terzo canone è quello del *metodo combinato di accordo e differenza*, caratterizzato dalla compresenza delle circostanze descritte nei primi due metodi. Un elemento può essere assunto come causa di un fenomeno se il fenomeno si verifica in più circostanze diverse fra loro, ma in cui l'elemento considerato è presente, mentre non si verifica in circostanze del tutto simili ma in cui l'elemento considerato è assente. Il quarto è il *metodo della variazione concomitante*, in base al quale se ad una variazione di un fenomeno segue sistematicamente la variazione di un secondo fenomeno, il primo può essere considerato causa e il secondo effetto. Il quinto è il *metodo dei residui*, che si applica quando cause ed effetti multipli sono presenti; dopo che una parte delle cause e degli effetti sono stati determinati, si può inferire che ciò che resta degli effetti sarà determinato dai fattori causali rimanenti².

Definito in questi termini, il metodo scientifico per determinare le relazioni di causalità si rivela però insufficiente a causa di alcuni limiti del ragionamento di Mill. I cinque canoni sono infatti basati sull'assunto che il mondo che ci circonda e l'apparente caos che lo caratterizza siano in realtà costituiti da *regolarità* empiriche assolute che possono essere scoperte e rivelate dallo scienziato, che le cause siano collegate in modo deterministico agli effetti. In altri termini, i canoni di Mill si basano sull'idea che nello studio di un fenomeno sia possibile identificare ed analizzare tutte le cause possibili, così da arrivare ad una spiegazione quasi meccanicistica dei fenomeni. Il problema sta nel fatto che la realtà è più complessa di quanto ipotizzato da Mill e trovare delle regolarità empiriche con le caratteristiche richieste dai suoi canoni è impossibile, specialmente nelle scienze sociali. Proviamo a chiarire il problema con un esempio ripreso da Thye (2007): immaginiamo che un gruppo di amici vada a cena in un ristorante e che il giorno dopo due di essi, che chiameremo Piero e Paola presentino i sintomi di un'intossicazione alimentare. È assai

² I primi 3 metodi di Mill richiamano i concetti aristotelici di condizione necessaria e sufficiente. In particolare, il metodo dell'accordo rimanda alla condizione necessaria: l'effetto y si verifica sempre in presenza dell'antecedente x , mentre il metodo della differenza corrisponde alla condizione sufficiente: l'effetto y non si verifica in assenza dell'antecedente x . Il quarto metodo invece richiama il problema delle relazioni spurie menzionato all'inizio del capitolo. Le causalità spurie sono state escluse dal ragionamento di Mill assumendo che la variazione dei due fenomeni considerati non sia concomitante, cioè avvenga in modo regolare e prevedibile, ma non simultanea.

probabile che il giorno dopo i due amici, sofferenti e preoccupati, telefonino ai commensali per verificare se qualcuno dei loro amici accusi gli stessi sintomi, per tentare di determinarne le cause. In tali circostanze, seguendo inconsapevolmente i precetti di Mill, Piero scoprirà che anche Paola sta male e che, pur avendo ordinato un menu completamente diverso dal suo, anche Paola ha mangiato le cozze (metodo dell'accordo). Inoltre, telefonando agli altri amici Piero e Paola potrebbero scoprire che nessun altro dei commensali ha accusato gli stessi problemi ed anche che, benché alcuni amici abbiano ordinato un menu simile al loro, nessun altro ha mangiato le cozze (metodo della differenza). Date le circostanze, Mill concluderebbe inequivocabilmente che la causa dell'intossicazione alimentare debba essere l'aver mangiato le cozze e, nei termini dell'esempio, questa conclusione potrebbe anche essere corretta. In linea di principio, però, questa non è l'unica spiegazione possibile, il ragionamento così come è stato presentato non permette di escludere fattori causali alternativi. Per esempio, potrebbe essere successo che casualmente Piero e Paola abbiano entrambi mangiato qualcosa nel pasto precedente la cena, e che questa sia la causa che ha determinato il malessere successivo. Oppure la causa potrebbe essere di natura completamente diversa, Piero e Paola potrebbero aver incontrato nel pomeriggio precedente un altro amico, che poi non ha partecipato alla cena, e da quest'ultimo aver contratto un virus intestinale. Oppure ancora potrebbe essere che, magari a loro insaputa, Piero e Paola condividano un'intolleranza per uno degli ingredienti utilizzati per la preparazione delle cozze e che quindi la vera causa dei sintomi da intossicazione sia una combinazione unica della loro intolleranza e del cibo consumato (quello che in statistica si chiama un effetto di interazione).

Nonostante questi limiti, il metodo utilizzato nelle scienze sperimentali, deve molto al lavoro di Mill, in particolare al secondo canone, il metodo della differenza. Il disegno di un esperimento, nella sua forma più semplice, consiste infatti nella creazione di due situazioni per quanto possibile identiche, tranne che per un solo elemento, e nell'osservazione di un fenomeno in queste due condizioni. Il fenomeno è l'*explanandum*, l'effetto, la variabile dipendente dello studio. L'elemento differente è tipicamente l'*explanans*, il fattore che il ricercatore ha ipotizzato possa essere la causa del fenomeno oggetto di studio. Pertanto l'elemento differente è l'oggetto concreto della manipolazione sperimentale, è il fattore di diversificazione che il ricercatore introduce artificialmente in una fra due situazioni altrimenti identiche. La condizione nella quale viene introdotta la manipolazione sperimentale viene comunemente indicata come *condizione (o trattamento o gruppo) sperimentale*, mentre la seconda, quella in cui la manipolazione è assente, viene indicata come *condizione (o trattamento o gruppo) di controllo*. Nei termini di Mill questa procedura dovrebbe dar luogo a risultati del tipo: se il fattore x è presente si verifica il fenomeno y , se il fattore x è assente il fenomeno y non si verifica. Nella pratica invece i risultati ottenuti saranno al massimo *probabilistici*: l'evento y si verifica più spesso in presenza del fattore x e raramente quando il fattore x è assente.

Lo stesso Mill si rese conto che i suoi canoni permettevano al massimo di identificare cause probabilistiche, ma egli considerava queste come un passaggio intermedio nel processo di identificazione di relazioni causali universali. Invece, fu il biologo e matematico inglese Ronald Fischer (1935, 1956) a porre le relazioni causali di tipo probabilistico co-

me obbiettivo finale della ricerca sperimentale, utilizzando la teoria della probabilità e l'inferenza statistica. Gli esperimenti basati sul metodo della differenza di Mill erano già comuni nell'800, specialmente nel campo della biologia e della medicina. Fischer (1935) descrisse un esperimento condotto da Darwin sui vantaggi relativi all'impollinazione incrociata in confronto all'autoimpollinazione. Confrontando due gruppi di piante, uno ottenuto da semi prodotti tramite autoimpollinazione e l'altro da semi ottenuti per impollinazione incrociata, Darwin osservò che le piante del secondo gruppo erano più grosse, come lui aveva ipotizzato sulla base della sua teoria dell'evoluzione mediante selezione naturale. Nondimeno, Darwin non fu in grado, nemmeno con l'aiuto dei migliori matematici della sua epoca, di stabilire se la differenza osservata fa i due campioni fosse sufficiente a corroborare la sua ipotesi o non potesse invece essere semplicemente il prodotto del caso o di qualche altro possibile fattore non considerato.

La soluzione al problema delle possibili cause alternative non considerate proposta da Fischer è relativamente semplice e consiste nell'assegnare in modo casuale gli individui che partecipano all'esperimento alle due condizioni. Per illustrare la metodologia proposta da Fischer utilizzeremo, come esempio, il celebre esperimento di Asch (1951) sul conformismo, descritto nel Riquadro 5.1. Asch assegnò in modo casuale i partecipanti al suo esperimento a due gruppi di numerosità simile, un gruppo fu sottoposto al trattamento sperimentale, l'altro al trattamento di controllo. Trattandosi di un gruppo di studenti universitari maschi, dei partecipanti a questo esperimento sappiamo solo che erano uniformi rispetto ad età e genere, ma non abbiamo nessuna informazione relativa a qualsiasi altro fattore potenzialmente significativo nel determinare risposte difformi allo stimolo utilizzato nell'esperimento. Potremmo non sapere nulla della loro famiglia d'origine, della loro estrazione sociale, della loro personalità, della loro storia di vita ecc., e alcuni di questi fattori potrebbero essere associati ad una diversa propensione individuale a cedere alla pressione del gruppo e fornire risposte conformi. Naturalmente alcuni di questi fattori, tipicamente quelli maggiormente rilevanti da un punto di vista teorico, potrebbero essere tenuti sotto controllo, raccogliendo le informazioni necessarie per mezzo di un questionario somministrato ai soggetti durante l'esperimento. Il problema però permane perché la lista dei fattori potenzialmente rilevanti è virtualmente infinita. L'assegnazione casuale risolve il problema perché assicura che tutti i fattori potenzialmente rilevanti (noti e ignoti) siano ugualmente rappresentati nei due gruppi dal momento che, per ognuna delle caratteristiche rilevanti, individui con caratteristiche diverse hanno la stessa probabilità (50%) di finire in un gruppo o nell'altro.

Il campione di studenti utilizzato da Asch proveniva dall'universo costituito da tutti i possibili campioni della stessa numerosità, estraibili dalla popolazione di riferimento, cioè la popolazione degli studenti maschi dello Swarthmore College, dove fu condotto l'esperimento. La variabile effetto rilevata da Asch fu il tasso di errore nelle risposte fornite dai partecipanti. Poiché l'assegnazione casuale garantisce l'uguaglianza fra i campioni su tutti i fattori rilevanti, in assenza di stimoli volti a produrre il conformismo, il tasso di errore medio dovrebbe dipendere soltanto dalla variazione casuale fra i tassi medi di errore di tutti i campioni estraibili da quella popolazione, compreso quello che venne assegnato al trattamento sperimentale.

Riquadro 5.1

Gli esperimenti sul conformismo di Asch

Negli anni 50 lo psicologo sociale Solomon Asch condusse una serie di celebri esperimenti per studiare le condizioni sociali in cui gli individui tendono a cedere alla pressione del gruppo e uniformarsi alle opinioni di quest'ultimo quando esse siano palesemente contrarie alla realtà fattuale. Nel primo di questi esperimenti Asch divise gli 87 studenti maschi che parteciparono in due gruppi. Gli studenti assegnati alla condizione sperimentale venivano fatti sedere, uno alla volta, in un semicerchio insieme ad altri 7 studenti, che erano in realtà complici del ricercatore istruiti a fornire risposte errate ad alcune di quelle che Asch chiamò «prove critiche». Il compito dei partecipanti era di rispondere ad una serie di domande relative al confronto fra alcune linee che venivano loro mostrate. Ai partecipanti vennero mostrate 18 coppie di carte come quelle rappresentate nella figura 5.1, dopo di che venne loro chiesto di indicare quale dei 3 segmenti rappresentati nella carta di destra fosse di lunghezza uguale a quello rappresentato nella carta di sinistra. In ciascuna coppia la risposta corretta era sempre ovvia.

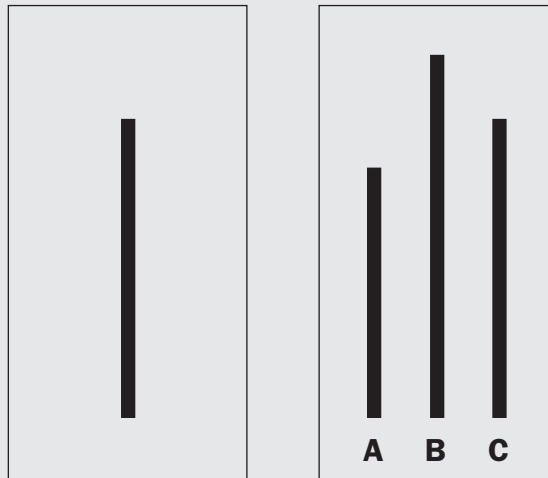


Figura 5.1 Lo stimolo sperimentale utilizzato da Asch.

Nel trattamento sperimentale, in 12 delle 18 prove, i complici, che venivano sollecitati a rispondere prima del partecipante reale, fornivano sistematicamente una risposta errata. Nel trattamento di controllo tutti i partecipanti rispondevano in privato, scrivendo la propria risposta su un foglio. In questo modo le due condizioni venivano mantenute identiche ad eccezione della manipolazione sperimentale: l'esposizione all'opinione del gruppo. I 50 soggetti nel gruppo sperimentale commisero in media 3,84 errori, con un tasso medio di errore pari al 32%, quelli nel gruppo di controllo 0,08 errori (0,007%).

Di conseguenza, è piuttosto improbabile che la differenza macroscopica rilevata da Asch, fra i tassi medi di errore dei due gruppi, possa essere attribuita a una variazione nel tasso medio di errore della popolazione indipendente dalla manipolazione, cioè essenzialmente al caso. Quanto improbabile? Fischer fornì la risposta a questa domanda grazie al contributo di un altro matematico inglese, William S. Gossett, l'ideatore del t test³, che inventò un metodo per utilizzare la varianza del campione, al posto di quella della popolazione, che è ignota, per stimare l'errore standard⁴. L'analisi dei dati sperimentali permette così di stimare la probabilità che una data differenza, per esempio nel caso di Asch fra i tassi di errore delle due condizioni, si verifichi per cause accidentali, cioè in assenza di un effetto determinato dalla manipolazione sperimentale. Convenzionalmente, si conclude che esiste una relazione causale fra la manipolazione e la differenza osservata se la probabilità che quest'ultima sia determinata dal caso è inferiore al 5%, che viene considerato la soglia di significatività statistica. Applicando questo metodo ai dati dell'esperimento di Darwin citato in precedenza, Fischer poté concludere che l'ipotesi di Darwin era corretta, in quanto la differenza nelle dimensioni delle piante dei due gruppi era statisticamente significativa.

Benché il metodo proposto da Fischer sia tuttora universalmente utilizzato nella ricerca sperimentale, questo approccio è stato fortemente criticato da alcuni ricercatori. In particolare Bernard Cohen (1997) criticò il metodo sperimentale proponendo una «onorevole sepoltura» per Mill e Fischer, sulla base di due ragionamenti, il primo basato sul ruolo dei potenziali fattori causali ignorati e il secondo sulla relazione fra assegnazione casuale dei soggetti alle condizioni sperimentali e numerosità del campione. Innanzitutto Cohen sostenne che poiché il numero di potenziali fattori causali ignoti è virtualmente infinito, la probabilità che i due sottocampioni di un esperimento finiscano per differire su almeno uno di questi si avvicina a 1. In secondo luogo, secondo Cohen, sebbene sia vero che l'assegnazione causale ha l'effetto di rendere i due sottocampioni uguali, questo avviene soltanto se i due sottocampioni hanno numerosità infinita. In risposta a queste critiche, altri (Thye 2007, Willer e Walker 2007) hanno sostenuto che, benché sia vero che la condizione sperimentale e quella di controllo possano differire per alcune caratteristiche non osservate, la maggior parte delle caratteristiche non osservate sono irrilevanti rispetto

³ Gossett sviluppò la statistica t mentre confrontava la qualità di piccoli campioni di orzo utilizzati per produrre la birra dalla Guinness, presso cui lavorava come chimico. Poiché la politica aziendale non permetteva alcuna pubblicazione ai propri dipendenti, Gossett pubblicò il suo lavoro con lo pseudonimo Student, con il quale viene solitamente indicata la distribuzione t , in suo onore (Student 1908).

⁴ La varianza è una misura della variabilità, o più tecnicamente dispersione, di una variabile, ottenuta come media degli scostamenti quadratici dal valore medio della variabile. Per scopi descrittivi viene più spesso riportata la deviazione standard che, essendo la radice quadrata della varianza, è intuitivamente interpretabile, nella stessa unità di misura della variabile, come scarto medio rispetto alla media. Infine, l'errore standard è una stima della variabilità e quindi dell'imprecisione di una stima statistica. Per esempio, solitamente si usa la media di un campione come stima della media di una popolazione, ma ogni campione di uguale numerosità può avere una media diversa, quindi il grado di precisione della stima dipenderà da quanto è variabile la media ottenibile da ciascuno dei campioni possibili. L'errore standard è una misura di questa variabilità.

al problema oggetto di studio, per esempio se in fisica si studia la caduta dei corpi materiali nel vuoto, il colore dei medesimi sarà irrilevante. Analogamente, non pare particolarmente importante sapere se gli studenti dei due gruppi che parteciparono all'esperimento di Asch differissero significativamente per peso o altezza. Inoltre, per il teorema del limite centrale, l'effetto aggregato dei fattori causali inosservati che differenziano le condizioni di un esperimento tende ad essere normalmente distribuito, e poiché alcuni di questi fattori produrranno effetti in un senso ed altri nel senso contrario, il loro effetto combinato sarà di produrre una distorsione non sistematica, generalmente detta rumore (noise), nella misurazione della variabile dipendente, e questo rumore avrà una distribuzione normale con un valore atteso prossimo allo zero (Thye 2007).

In merito al secondo punto sollevato da Cohen (1997) occorre precisare che esiste una relazione fra la numerosità del campione e gli errori che si possono commettere nel processo di inferenza casuale. Gli errori che si possono commettere sono due: l'errore di primo tipo o falso positivo e l'errore di secondo tipo o falso negativo, riassunti nella tabella 5.1. Nel caso di un esperimento l'errore di primo tipo consiste nel concludere erroneamente che la manipolazione ha prodotto un effetto, quando invece l'effetto osservato è semplicemente il prodotto di una variazione casuale, per esempio dovuta alla scelta di un campione «insolito». L'errore di primo tipo, detto α , è quindi rappresentato dalla soglia di significatività che si utilizza per valutare la differenza trovata nella variabile dipendente fra i due gruppi dell'esperimento e, come detto in precedenza, convenzionalmente nelle scienze sociali si accetta un rischio di commettere questo tipo di errore non superiore al 5%. L'errore di secondo tipo invece consiste nella situazione opposta, cioè si commette questo errore quando si conclude che la manipolazione sperimentale non ha prodotto alcun effetto sul fenomeno oggetto di studio, mentre in realtà l'effetto esiste. Questo errore viene solitamente indicato con la lettera greca β ed è direttamente collegato alla potenza statistica del test, che è definita come $1-\beta$, cioè in altri termini la potenza di un test è definita come la probabilità di non commettere un errore di secondo tipo, cioè concludere che l'effetto è assente quando invece è presente⁵.

Tabella 5.1 Tipi di errore dei test statistici

Ipotesi di ricerca	Supportata dal test	Non supportata dal test
Valida	Decisione corretta: vero positivo	Errore di secondo tipo: falso negativo
Sbagliata	Errore di primo tipo: falso positivo	Decisione corretta: vero negativo

In base alla seconda critica mossa da Cohen al metodo sperimentale, la numerosità necessariamente finita del campione potrebbe produrre una differenza nella variabile dipendente determinata da fattori inosservati anziché dalla manipolazione sperimentale,

⁵ La Potenza statistica del test è molto importante particolarmente in medicina, per i test diagnostici, dove ovviamente è fondamentale che i soggetti malati vengano correttamente identificati come tali dal test.

cioè di fatto un errore del primo tipo. Per contro, sebbene sia vero che il rischio ipotizzato da Cohen diminuisce asintoticamente all'aumentare della numerosità del campione, il problema non è semplicemente risolvibile utilizzando campioni di dimensioni molto grandi. Infatti, poiché la probabilità di trovare una differenza statisticamente significativa nei risultati di un dato esperimento aumenta all'aumentare della numerosità del campione, utilizzando campioni troppo numerosi di fatto si aumenta il rischio di commettere l'errore di primo tipo. Per campioni molto grandi anche una differenza microscopica risulterà statisticamente significativa. Per questo motivo, accanto alla significatività statistica è importante tenere in considerazione anche la magnitudine dell'effetto trovato. Negli esperimenti più famosi, come per esempio quello di Asch, si ha il cosiddetto «effetto mazza» (Willer e Walker 2007): l'effetto trovato è talmente grande che sia l'obiezione di Cohen, che qualsiasi considerazione sulla dimensione del campione, e persino la significatività dei test statistici perdono rilevanza⁶.

Per valutare la magnitudine dell'effetto di un esperimento gli psicologi sono soliti utilizzare degli indicatori quantitativi noti come *dimensioni dell'effetto* (effect size). Nel caso di un esperimento come quello di Asch i cui risultati possono essere analizzati per mezzo di un semplice *t* test, la misura utilizzata è la *d* di Cohen, calcolata come la differenza fra i valori medi delle due condizioni divisa per la deviazione standard (Cohen 1992). Convenzionalmente un effetto viene considerato debole se $d = 0,2$, medio se $d = 0,5$ e forte se $d = 0,8$ ⁷. Per esempio, la dimensione dell'effetto trovato nell'esperimento di Asch è molto forte: $d = 1,5$. Nella fase di progettazione di un esperimento, stabilito il tipo di test con cui verrà testata l'ipotesi, è possibile scegliere la soglia di probabilità che si vuole accettare per il rischio di commettere un errore di primo tipo e la potenza statistica che si desidera che il test abbia. Dopo aver fissato questi due parametri, stimando un valore minimo per la dimensione dell'effetto che desideriamo che il nostro test sia in grado di cogliere, è possibile determinare la numerosità campionaria necessaria effettuando una analisi della potenza statistica (power analysis) come nell'esempio riportato nella figura 5.1⁸.

L'esempio riportato nel grafico si riferisce ad un semplice esperimento con due condizioni di numerosità uguale (tasso di allocazione $N1/N2 = 1$), i cui risultati verranno analizzati per mezzo di un *t* test ad una coda con una soglia di significatività del 5% e per un effetto atteso di dimensione media, vale a dire una differenza attesa fra le medie dei due gruppi pari ad almeno mezza deviazione standard ($d = 0,5$)⁹. Per esempio il grafico

⁶ A questo proposito, nella *review* di un articolo che presentava i risultati di un esperimento, corredati da analisi statistiche a suo avviso eccessivamente sofisticate, un *referee* anonimo ha scritto: «in un esperimento ben congegnato e riuscito l'unico test veramente necessario è il test del trauma interoculare: la verità deve colpire il lettore in mezzo agli occhi».

⁷ L'indice *d* di Cohen può essere trasformato in una misura percentuale di efficacia del trattamento. Per esempio $d = 0,2$ indica che il trattamento sperimentale ha un'efficacia del 58% nel senso che il 58% dei partecipanti del gruppo di controllo ha un punteggio sulla variabile dipendente inferiore al valore medio del gruppo sperimentale, per $d = 0,5$ l'efficacia è del 69% e per $d = 0,8$ l'efficacia è del 79%.

⁸ La figura 5.1 è stata ottenuta utilizzando il programma di power analysis G*Power (Faul et al. 2007).

⁹ Un test a due code si usa per verificare l'ipotesi che un effetto non sia nullo (quindi che sia positivo o

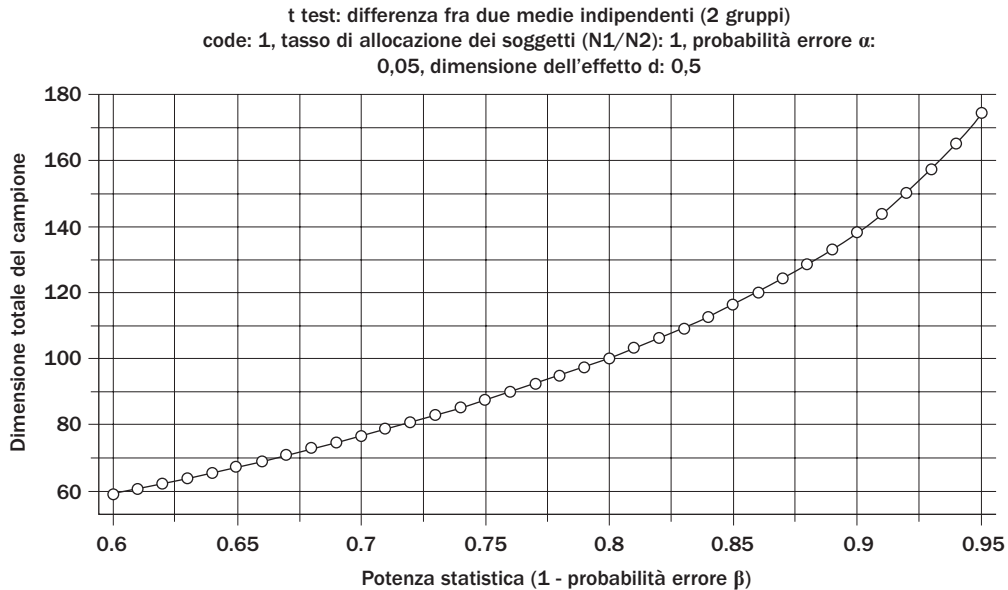


Figura 5.1 Power Analysis: dimensione campionaria in funzione della potenza statistica del test.

permette di stabilire che, se si desidera una potenza statistica pari al 95%, cioè una probabilità non superiore al 5% di ottenere un falso negativo, e quindi di non rilevare un effetto che è in realtà presente, è necessario un campione di circa 175 partecipanti.

Naturalmente se l'effetto atteso fosse di dimensione maggiore, la curva si sposterebbe verso sinistra e il campione richiesto sarebbe di dimensioni minori, per esempio per $d = 0,8$, sarebbero sufficienti 68 partecipanti. Analogamente, è possibile calcolare, a posteriori, la *sensibilità* di un esperimento, cioè la dimensione dell'effetto minimo rilevabile dato il numero di soggetti sperimentali utilizzati e la potenza statistica desiderata. Per esempio, l'esperimento di Asch aveva il 95% di probabilità di fornire un risultato statisticamente significativo se la differenza fra il numero medio di errori nei due gruppi fosse stata almeno pari a 0,7 deviazioni standard ($d = 0,7$).

L'assegnazione casuale dei soggetti sperimentali alle condizioni solitamente non è, in senso stretto, possibile nel caso degli esperimenti di campo, ma per il resto l'approccio adottato rispecchia completamente quello degli esperimenti di laboratorio e i ricercatori utilizzano tutti gli accorgimenti per rendere le condizioni sperimentali il più possibile simili, evitando possibili *bias* causati dal controllo ridotto sulla selezione dei partecipanti,

negativo); con un test ad una coda, invece, si verifica solo l'ipotesi che un effetto sia positivo (o solo che sia negativo). Pertanto, dato il medesimo numero di partecipanti, il test a una coda ha una potenza statistica maggiore di quello a due code. L'uso del test a una coda è giustificato dalle ipotesi teoriche, per esempio nell'esperimento di Asch ci si aspetta che il numero medio di errori nelle risposte sia maggiore (e non minore) nel gruppo sperimentale rispetto al gruppo di controllo.

come illustrato da una serie di esperimenti di campo sulla «diffusione del disordine» condotti da un gruppo di sociologi olandesi nella città di Groningen (Keizer, Lindenberg e Steg 2006). Uno di questi esperimenti è descritto nel Riquadro 5.2.

Riquadro 5.2

La diffusione del disordine

Nel 2008 la rivista *Science* pubblicò un articolo che conteneva sei esperimenti di campo condotti da due psicologi sociali, Kees Keizer e Linda Steg, e un sociologo, Siegwart Lindenberg dell'università di Groningen. Lo scopo di questi esperimenti era di testare la famosa teoria delle «finestra rotte» (Wilson e Killing 1982), in base alla quale la presenza di disordine urbano sotto forma di graffiti, rifiuti e segni di atti vandalici, come appunto finestre con i vetri rotti, faciliterebbe la proliferazione di comportamenti antisociali e microcriminalità. Il meccanismo postulato da questa teoria implica che la violazione di una norma sociale produca una sorta di effetto di contagio, che può portare ad ulteriori infrazioni di altre norme. Le norme sociali possono essere divise in norme prescrittive e descrittive (Cialdini et al. 1990), le prime comprendono sia le norme formali del diritto, sia quelle informali in base alle quali un dato comportamento è associato ad approvazione o disapprovazione sociale. Le norme descrittive sono invece quelle che si inferiscono semplicemente osservando il contesto in cui gli individui si trovano e il comportamento degli altri attori presenti, ricavandone informazioni in merito a quale forma di comportamento possa ritenersi appropriata. Quando i due tipi di norme sono disallineati, per esempio nel caso in cui la norma descrittiva non rifletta le indicazioni della norma prescrittiva, la seconda tende a perdere di efficacia. Keizer e colleghi chiamarono questo fenomeno «effetto Cialdini» in onore del ricercatore che per primo lo teorizzò.

Nel primo dei sei esperimenti presentati il contesto scelto fu un parcheggio a rastrelliera per biciclette collocato in un vicolo della città di Groningen, accanto all'ingresso di un supermercato. In questo esperimento i ricercatori olandesi manipolarono la coerenza fra norme prescrittive e descrittive. La norma descrittiva fu implementata collocando un cartello di divieto che proibiva esplicitamente i graffiti sul muro adiacente al parcheggio delle bici. Nella condizione di controllo (norma descrittiva allineata alla norma prescrittiva) i ricercatori tinteggiarono il muro, nella notte precedente il giorno dell'esperimento, per renderlo immacolato; nella condizione sperimentale, i ricercatori collocarono una serie di graffiti sul muro, proprio accanto al cartello di divieto, nella notte precedente il giorno dell'esperimento. Per operativizzare la variabile dipendente, invece, i ricercatori collocarono, in entrambe le condizioni sperimentali, dei volantini pubblicitari sui manubri delle biciclette parcheggiate nella rastrelliera accanto al muro (dato che la bicicletta è il mezzo di trasporto più comune nei Paesi Bassi, non è insolito che i volantini pubblicitari vengano collocati sulle biciclette). Nel caso specifico i volantini contenevano gli auguri di buone vacanze da parte di un inesistente negozio di articoli sportivi. Successivamente, i ricercatori osservarono il comportamento di 140 individui adulti (70 per condizione) nel momento in cui ritiravano la propria bici dal parcheggio, e contarono quanti fra questi gettavano il volantino per terra (ovviamente non era stato collocato alcun cestino per l'immondizia nei paraggi del parcheggio). Le strade di Groningen sono generalmente pulite e adeguatamente controllate dalla polizia, gettare rifiuti per terra comporta un'infrazione che può essere punita con una sanzione economica. Il 33% dei soggetti osservati nella condizione di controllo gettò il volantino per terra, nella condizione sperimentale la percentuale rilevata fu del

69%. Negli altri cinque esperimenti di campo i ricercatori utilizzarono situazioni relativamente simili ma variando il tipo di norma prescrittiva infranta (formale versus informale) e il tipo di infrazione osservata come variabile dipendente, utilizzando tipi di infrazione relativamente innocenti ed altri più gravi, fino ad arrivare al furto. In tutti e sei gli esperimenti i risultati furono analoghi: le infrazioni furono significativamente più frequenti nei contesti in cui una qualche norma prescrittiva era palesemente violata.

Nei sei esperimenti condotti da Keizer e colleghi i soggetti partecipanti non furono scelti dai ricercatori, che si limitarono invece ad osservare il comportamento degli individui che si trovarono causalmente ad attraversare il luogo scelto come setting per l'esperimento. Nel caso dell'esperimento descritto nel Riquadro 5.2 i partecipanti, completamente ignari, furono coloro che parcheggiarono la propria bici nel luogo scelto dai ricercatori, nella maggior parte dei casi probabilmente per recarsi al supermercato accanto al quale era collocato il parcheggio a rastrelliera. Come è piuttosto comune per gli articoli pubblicati dalla rivista *Science*, che hanno un formato molto compatto, l'articolo è accompagnato da un secondo documento (scaricabile in un file separato dal sito della rivista) che contiene l'appendice metodologica. Spesso, in particolar modo per gli articoli che trattano i risultati di ricerche sperimentali, l'appendice metodologica è più lunga dell'articolo stesso e contiene una serie di informazioni supplementari relative alla procedura utilizzata. Nell'appendice metodologica dell'articolo di Keizer e colleghi sono descritti gli accorgimenti presi dai ricercatori per riprodurre le condizioni tipiche di un laboratorio e rendere le due condizioni dell'esperimento virtualmente identiche. Per esempio i soggetti contattati come partecipanti furono solo quelli che nel periodo di osservazione parcheggiarono la propria bici entro un raggio di tre metri dal muro su cui erano stati collocati i graffiti, in modo da assicurarsi che la contraddizione (o coerenza) fra norma descrittiva e norma prescrittiva, che costituisce l'elemento di diversificazione delle due condizioni e la variabile indipendente dell'esperimento, fosse ugualmente visibile per tutti i partecipanti. Ciononostante i volantini furono collocati su tutte le biciclette parcheggiate nel vicolo per evitare di indurre un possibile senso di frustrazione nei partecipanti, dovuto all'osservazione che le altre biciclette parcheggiate un po' più in là non avevano ricevuto il volantino. Tutti i volantini gettati per terra vennero immediatamente raccolti dai ricercatori per evitare che i risultati potessero essere influenzati dalla presenza di una ulteriore infrazione normativa, in questo caso relativa alla variabile dipendente, mediante un effetto di imitazione. Le due condizioni sperimentali furono infine rese ulteriormente comparabili effettuando gli esperimenti alla stessa ora (fra le 13 e le 17) di un giorno ferialo, e in condizioni meteorologiche equivalenti (cielo sereno).

5.2 Il disegno sperimentale

Il modello più semplice di disegno sperimentale è quello che si ritrova comunemente negli esperimenti condotti in ambito medico dove il problema solitamente consiste nel determinare se un dato farmaco sia efficace nella cura di una specifica patologia. Il disegno utilizzato per questo tipo di esperimento è simile a quello utilizzato da Asch: 1) si estrae un campione di numerosità adeguata dalla popolazione di riferimento, cioè quella dei pazienti affetti dalla patologia al cui trattamento il farmaco è destinato; 2) i pazienti selezionati vengono assegnati in modo causale ad una delle due condizioni e 3) ai soggetti assegnati al gruppo sperimentale viene somministrato il farmaco, mentre a quelli del gruppo di controllo non viene somministrato nulla, o più spesso viene somministrato un placebo, per poter valutare l'effetto del farmaco al netto dell'effetto placebo¹⁰.

Per quanto l'approccio sia simile, nelle scienze sociali il disegno di un buon esperimento è tipicamente un'operazione più complessa. Per esempio, spesso lo stimolo sperimentale che si deve somministrare non è per sua natura dicotomico, ma può variare in modo continuo rispetto all'intensità. Inoltre la misurazione delle variabili effetto studiate dagli scienziati sociali è talvolta problematica in funzione del grado di affidabilità della misura utilizzata¹¹. La scarsa affidabilità di una misura si manifesta come componente random dell'errore che influenza la misura rilevata. L'effetto di una bassa affidabilità dello strumento di misura utilizzato per la variabile dipendente influenza i risultati dei test statistici rendendoli più conservativi, cioè nel caso di un esperimento, deprimendo la stima degli effetti causali studiati¹². Un secondo ordine di problemi è quello che riguarda la validità esterna ed interna dei risultati di uno studio scientifico. Nel caso degli esperimenti di laboratorio, la validità esterna si riferisce al grado in cui i risultati ottenuti sono validi al di fuori del contesto artificiale del laboratorio, e cioè in che misura i fenomeni osservati in laboratorio siano rappresentazioni verosimili di fenomeni analoghi che si verificano nella vita normale delle persone. Il problema della validità esterna è assimilabile al problema della generalizzabilità alla popolazione di riferimento dei risultati ottenuti mediante dati di survey condotte su un campione di quella popolazione, ma a differenza del caso delle survey, per gli esperimenti non si tratta di un problema di rappresentatività statistica del campione utilizzato, ma piuttosto del grado di realismo o verosimiglianza della situazione studiata nel contesto sperimentale¹³. Il problema della validità interna si

¹⁰ L'effetto placebo è quello che si ha quando i pazienti mostrano segni di miglioramento per il solo fatto di essere sottoposti ad un trattamento, anche nel caso in cui il trattamento sia completamente privo di proprietà curative o farmacologiche. Questo effetto si manifesta a condizione che i pazienti ripongano fiducia nel trattamento ricevuto.

¹¹ L'affidabilità o attendibilità di una misura statistica si riferisce al grado in cui lo strumento di misura utilizzato produce risultati simili in circostanze simili. Uno strumento attendibile non è necessariamente valido, perché la validità dipende dallo scopo per cui si usa una data misura. La distorsione della misura prodotta da uno strumento poco attendibile è distribuita in modo casuale, mentre uno strumento non valido produce una distorsione sistematica.

¹² Per una rassegna dei problemi legati all'affidabilità nella ricerca sperimentale, si veda Thye (2000).

¹³ Durante una lezione sul metodo sperimentale tenuta nel 2002 a un corso estivo in economia sperimentale-

riferisce invece ad una questione più sottile. La validità interna «si riferisce al grado in cui un metodo è in grado di stabilire una relazione di causa ed effetto» (Thye 2007: 72). Seguendo Thye (2007), ora illustreremo le principali fonti di disturbo della validità interna utilizzando un esempio.

Immaginiamo di voler disegnare un esperimento per valutare l'efficacia di un programma di intervento applicato in una o più scuole medie inferiori, allo scopo di prevenire o mitigare l'assunzione di comportamenti a rischio da parte degli adolescenti. Per comportamenti a rischio si intende il fumo, il consumo di alcolici, la tendenza a comportamenti violenti e atti di delinquenza minorile quali il taccheggio o il vandalismo. Questo genere di interventi mira in ultima analisi a prevenire la dispersione scolastica, che è notoriamente associata all'assunzione precoce di comportamenti a rischio. Immaginiamo di partire da una teoria in base alla quale la tendenza verso i comportamenti a rischio può essere attenuata favorendo il coinvolgimento degli studenti in attività di gruppo, che mirino a rafforzare l'integrazione sociale degli adolescenti all'interno della scuola, facilitando l'adozione di norme che premino il perseguimento di buoni risultati scolastici e scorraggino la devianza minorile. Il programma in questione potrebbe essere costituito da un certo numero di ore da dedicare ad attività di gruppo volte a stimolare l'integrazione degli studenti. Un primo approccio potrebbe essere quello di implementare il programma, che chiameremo X , in alcune classi e successivamente misurare la propensione media ai comportamenti a rischio degli studenti coinvolti, che chiameremo O . Questo approccio può essere rappresentato schematicamente in questo modo:

$$X \rightarrow O$$

Questo tipo di esperimento, definito «one-shot post hoc» (Thye 2007), non permette di trarre alcuna conclusione relativa all'efficacia del programma di prevenzione, perché manca una condizione di controllo o un qualsiasi termine di paragone con cui confrontare il grado di propensione media al rischio osservato in O . L'unico termine di paragone possibile potrebbe essere costituito dalla propensione media al comportamento a rischio della popolazione di riferimento, se fosse nota, ma possiamo realisticamente assumere che quest'ultima non sia nota. Inoltre, seppure la propensione media della popolazione fosse nota, il confronto sarebbe comunque reso fallace da una serie di elementi di disturbo che illustreremo più avanti. Una soluzione più efficace potrebbe essere quella di misurare la propensione media ai comportamenti a rischio prima di implementare il programma di

le, l'economista Simon Gächter motivò l'utilizzo di campioni di studenti universitari come soggetti negli esperimenti di laboratorio dicendo: «il metodo sperimentale ha una lunga tradizione in medicina e farmacologia, molti degli esperimenti effettuati da medici e biologi utilizzano i topi come soggetti sperimentali. La validità esterna di questi esperimenti non dipende dalla rappresentatività statistica, ma solo dal fatto che l'organismo dei topi da laboratorio può essere considerato una *approssimazione sufficiente* dell'organismo degli esseri umani. Analogamente, nelle scienze sociali sperimentali si utilizzano gruppi di studenti perché si crede che essi siano una approssimazione sufficiente degli esseri umani.»

prevenzione (O_1) e poi una seconda volta dopo la conclusione dell'intervento (O_2). Questo secondo approccio può essere rappresentato come segue:

$$O_1 \quad \rightarrow \quad X \quad \rightarrow \quad O_2$$

Con questo secondo approccio il ricercatore potrebbe confrontare la misura osservata dopo l'attuazione del programma con quella registrata in precedenza, e presumendo che il programma sia utile o il ricercatore piuttosto fortunato, rilevare una diminuzione nella media dei punteggi osservati. Utilizzando un esempio simile, Thye (2007) presenta una lista di nove possibili fonti di disturbo che riassumeremo di seguito.

Storia. La differenza osservata potrebbe essere stata causata da un elevato numero di fattori causali inosservati, indipendenti dal programma di intervento, per esempio, gli studenti potrebbero essere stati influenzati da una campagna nazionale di dissuasione dai comportamenti a rischio, che sia casualmente in corso durante il periodo di implementazione del programma. Questo tipo di disegno sperimentale rende impossibile separare gli effetti del programma di intervento che si vuole valutare da quelli di qualunque evento saliente che si verifichi nel medesimo periodo.

Maturazione. Anche in assenza di rilevanti eventi salienti nel periodo considerato, qualcosa accadrà in ogni caso agli studenti coinvolti: trascorrerà il tempo. Nel periodo considerato gli studenti potrebbero attraversare cambiamenti anche repentini dovuti semplicemente alla crescita e allo sviluppo. Questi cambiamenti spontanei potrebbero produrre una riduzione o un aumento delle attività a rischio indipendentemente dall'efficacia del programma di prevenzione portando nel primo caso ad una errata valutazione positiva del programma e nel secondo ad una errata valutazione negativa.

Selezione. Se la selezione delle classi e degli studenti coinvolti nel progetto avviene in maniera non casuale, gli individui selezionati potrebbero essere accomunati da caratteristiche specifiche che potrebbero renderli significativamente diversi dalla popolazione da cui provengono, e il grado di efficacia dell'intervento potrebbe essere influenzato da queste caratteristiche comuni ai soggetti coinvolti. Nei progetti di valutazione di efficacia di progetti di prevenzione, questo problema è molto comune in quanto le agenzie che finanziano interventi di questo tipo preferiscono che il denaro investito sia diretto agli individui maggiormente a rischio. Questa procedura di selezione però inficia la valutazione di efficacia, rendendo inadeguato qualunque termine di paragone si possa utilizzare per valutare i risultati ottenuti sul campione selezionato in questo modo.

Test. Il semplice fatto di misurare i comportamenti a rischio prima e dopo l'intervento in alcuni casi può influenzare l'efficacia dell'intervento stesso in due modi. Da una parte attraverso un effetto di sensibilizzazione, il grado di attenzione e coinvolgimento nelle attività del programma potrebbe cambiare come conseguenza dell'esposizione dei soggetti ad uno strumento di misurazione volto a rilevare la propensione ai comportamenti a rischio; dall'altra potrebbe esserci un effetto di « apprendimento » dello strumento utilizzato che, nel caso dell'esempio, potrebbe indurre i partecipanti a ricercare una qualche forma di coerenza fra le risposte fornite prima e dopo la somministrazione dello stimolo.

Regressione. Quando si usano misure ripetute, i punteggi estremi tendono a ridursi

con l'andare del tempo per effetto della cosiddetta *regressione verso la media*: i valori estremi tendono ad essere rari e quindi eventuali valori estremi osservati casualmente nella prima rilevazione tenderanno naturalmente a scomparire nella seconda, non per effetto dell'intervento, ma semplicemente per effetto del caso.

Affidabilità. Anche l'affidabilità delle misure utilizzate, di cui abbiamo già parlato in precedenza, oltre a costituire un problema di per sé, può essere fonte di disturbo della validità interna. Nella misura in cui lo strumento utilizzato avesse un basso grado di affidabilità, i valori ottenuti per mezzo di esso tenderanno a fluttuare naturalmente in rilevazioni successive nel tempo. Queste fluttuazioni tenderanno a disturbare la valutazione dell'efficacia del programma di intervento.

Mortalità sperimentale. Così come accade nelle ricerche che utilizzano survey longitudinali, anche negli esperimenti di questo tipo accade regolarmente che alcuni individui abbandonino l'esperimento, prima che il programma sia portato a termine e sia effettuata la seconda rilevazione. Se gli individui abbandonassero il programma in modo casuale (cioè se ognuno avesse la stessa probabilità di abbandonare l'esperimento), questo non costituirebbe alcun problema per la validità interna dello studio, purtroppo però spesso gli studenti che abbandonano sono concentrati su valori estremi della variabile dipendente, anziché distribuiti in modo casuale: i soggetti più propensi al rischio potrebbero abbandonare qualora il programma induca in loro un atteggiamento di rifiuto, quelli meno propensi potrebbero abbandonare perché ritengono il programma noioso o inutile.

Bias dello sperimentatore. Quest'ultima è forse la fonte di disturbo più nota e documentata nella letteratura sugli esperimenti di laboratorio e si riferisce al fatto che le aspettative del ricercatore tendono a influenzare, inconsapevolmente, i risultati ottenuti nell'esperimento. In un famoso esperimento su questo bias, Rosenthal (1966) istruì alcuni studenti a condurre un esperimento di laboratorio sulle capacità di apprendimento dei topi, che vennero divisi in due gruppi sulla base della loro presunta intelligenza (in realtà i due gruppi erano costituiti da topi perfettamente identici, collocati nei due gruppi in modo casuale). I risultati mostrarono che i topi che erano stati etichettati come intelligenti mostrarono un apprendimento superiore a quelli che erano stati definiti non intelligenti. Il bias dello sperimentatore può manifestarsi, come in questo caso, per effetto diretto dell'influenza del ricercatore, oppure talvolta perché i partecipanti intuiscono lo scopo dell'esperimento e tentano attivamente di assecondare il ricercatore. Gli esperimenti sull'obbedienza condotti da Milgram (descritti nel Riquadro 3.1) sono stati criticati non solo per la mancanza di un autentico gruppo di controllo ma anche per la presenza intrusiva, in quanto potenziale fonte di disturbo, del ricercatore sulla scena dell'esperimento (Willer and Walker 2007).

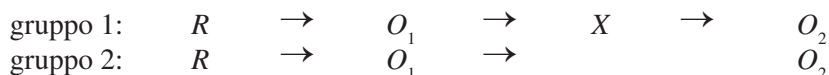
Il rimedio comunemente utilizzato per eliminare il bias dello sperimentatore è la cosiddetta procedura in *doppio cieco* (double blind). Mentre con il termine procedura «in cieco» (single blind) si intende la prassi mediante la quale ci si assicura che i soggetti partecipanti non conoscano le ipotesi di ricerca che vengono testate nell'esperimento, nella procedura in doppio cieco le ipotesi di ricerca vengono tenute nascoste anche al ricercatore che conduce l'esperimento (Thye 2007). Più in generale, i termini cieco e doppio cieco vengono utilizzati in letteratura anche per indicare, rispettivamente, che le intera-

zioni fra i soggetti partecipanti avvengono in forma anonima, e che sia le interazioni fra i partecipanti sia quelle fra i medesimi e il ricercatore che conduce l'esperimento avvengono in forma anonima. In questa seconda accezione per «forma anonima» si intende che le interazioni avvengono solo mediante computer, e che quindi ai soggetti viene impedito anche di vedersi (e quindi possibilmente riconoscersi); nel caso del doppio cieco anche il ricercatore interagisce con i partecipanti solo mediante computer. Questa completa anonimità ha due scopi: 1) prevenire la formazione di aspettative da parte dei partecipanti e quindi ridurre il bias dello sperimentatore e 2) garantire ai soggetti la piena libertà di agire come credono, entro i vincoli posti dall'esperimento stesso, senza doversi preoccupare degli effetti sociali o reputazionali delle proprie scelte. Come vedremo in seguito, proprio per questo secondo motivo la procedura in doppio cieco è particolarmente comune negli esperimenti degli economisti.

Secondo Thye (2007) tutte le altre fonti di disturbo della validità interna dell'esperimento possono essere eliminate utilizzando un disegno sperimentale che preveda la presenza di un autentico gruppo di controllo. Riprendendo l'esempio precedente questo può essere fatto in due modi. Si potrebbero assegnare alcune classi a due gruppi diversi e poi implementare il programma di prevenzione soltanto in uno dei due, creando così un trattamento sperimentale e uno di controllo. Possiamo rappresentare questa strategia nel modo seguente:



Questa strategia consente di eliminare le fonti di disturbo causate dalla storia e dalla regressione verso la media, in quanto i due gruppi vengono osservati nello stesso intervallo di tempo e quindi storia e regressione verso la media incidono sui risultati rilevati in modo analogo in entrambi i gruppi. Tuttavia, la mancanza di assegnazione casuale degli studenti ai due gruppi non permette di eliminare le altre fonti di disturbo. A causa della selezione non casuale i risultati dell'esperimento possono essere influenzati da tutte le caratteristiche personali sulla base delle quali gli individui nei due gruppi potrebbero essere sistematicamente diversi fra loro. Per molti degli psicologi sociali dediti alla ricerca sperimentale, incluso lo stesso Thye, soltanto un disegno che preveda 1) la presenza di un gruppo di controllo con il quale effettuare i confronti dei risultati ottenuti nei gruppi sperimentali e 2) l'assegnazione dei soggetti alle diverse condizioni su base perfettamente casuale, può essere considerato un «vero esperimento». Nel caso del nostro esempio, un disegno con queste caratteristiche può essere rappresentato come segue:



dove *R* (random) rappresenta appunto l'assegnazione dei soggetti alle due condizioni mediante una procedura casuale, che permette di rendere i due gruppi uguali all'origine e consente così di attribuire le eventuali differenze riscontrate nei risultati, in modo univo-

co, all'effetto della manipolazione sperimentale, cioè, nel caso del nostro esempio, al programma di prevenzione dei comportamenti a rischio somministrato agli studenti assegnati al gruppo 1.

Nonostante la presenza di almeno due condizioni, di cui una di controllo, e l'assegnazione casuale siano capisaldi della tradizione sperimentale, specialmente in psicologia sociale, esiste un tipo di disegno sperimentale mediante il quale possono essere effettuati esperimenti con una sola condizione e quindi senza assegnazione casuale dei soggetti, il cosiddetto disegno ABA (Meeker and Leik 2007). Il disegno ABA consiste di tre parti, contraddistinte dalla sequenza di tre lettere A, B, A. La prima parte (A) dell'esperimento corrisponde alla condizione di controllo, nella successiva parte (B) viene introdotto lo stimolo sperimentale che, terminato il periodo di osservazione, viene nuovamente eliminato nella terza ed ultima parte (A), che quindi ripristina le stesse condizioni osservate nella prima parte dell'esperimento. Questo disegno è adatto allo studio di quei fenomeni in cui l'effetto della manipolazione è visibile soltanto quando lo stimolo che lo produce è contemporaneamente presente, ma sparisce non appena lo stimolo viene rimosso, infatti l'uso di questo tipo di disegno è comune, negli esperimenti di psicologia sociale, soltanto fra i ricercatori della tradizione comportamentista, per i quali il comportamento umano è spiegabile come riflesso spontaneo a stimoli ricevuti dall'ambiente. In questo genere di esperimenti il gruppo di controllo non è necessario, perché l'effetto causale dello stimolo sperimentale sulla variabile dipendente viene dimostrato mediante l'annullamento dell'effetto determinato dalla scomparsa dello stimolo e, di conseguenza, possibili effetti spuri possono essere esclusi anche senza ricorrere all'assegnazione casuale. Questo tipo di manipolazione è nota anche con il nome di manipolazione *intra-soggetti* (within subjects), mentre la procedura classica con assegnazione dei soggetti a due condizioni distinte viene per contrasto chiamata manipolazione *inter-soggetti* (between subjects).

Oltre ai comportamentisti, la manipolazione intra-soggetti viene usata con minor riluttanza anche dagli economisti sperimentali, per esempio nel citatissimo esperimento di Fehr e Gächter (2002) sulla punizione altruistica. Riprendendo un'idea già presentata 16 anni prima dallo psicologo sociale Yamagishi (1986), peraltro senza citarlo, in questo esperimento i due economisti dimostrarono che il livello di cooperazione all'interno di piccoli gruppi varia drasticamente se ai soggetti sperimentali viene consentito di utilizzare un sistema sanzionatorio (al termine dell'interazione, che era costituita da un gioco del bene pubblico, descritto nel Riquadro 5.4, a ciascun partecipante veniva data la possibilità di spendere una piccola parte delle risorse accumulate per sanzionare uno degli altri partecipanti, riducendo i suoi profitti). In questo esperimento, i ricercatori utilizzarono una manipolazione intra-soggetti in cui il sistema sanzionatorio veniva aggiunto e tolto alla condizione di controllo. A differenza dei behavioristi però, Fehr e Gächter (e molti altri ricercatori che hanno utilizzato questo tipo di disegno) utilizzarono comunque due condizioni con assegnazione causale dei soggetti sperimentali: in una il sistema sanzionatorio era assente nella condizione iniziale (A) e veniva aggiunto nella seconda parte dell'esperimento (P), nell'altra il sistema sanzionatorio era presente nella parte iniziale (P) e rimosso nella seconda parte (A). Utilizzando lo stesso schema dell'esempio precedente potremmo rappresentare questo disegno così:



dove R rappresenta l'assegnazione casuale, e O_1 e O_2 rappresentano le osservazioni della variabile dipendente nelle due parti di ciascuna condizione. L'operazione di distinguere le due condizioni sperimentali di una manipolazione intra-soggetti, invertendo l'ordine di somministrazione dello stimolo, viene indicata con il termine di *controbilanciamento* (counterbalancing) e serve ad eliminare possibili *effetti dell'ordine*, cioè a verificare in che misura i risultati dell'esperimento siano sensibili all'ordine secondo il quale i soggetti vengono esposti allo stimolo.

Negli esperimenti illustrati fino a qui ci siamo soffermati su disegni sperimentali relativamente semplici, utilizzando solo esempi caratterizzati dalla manipolazione di un solo fattore causale. Nella pratica però molto spesso gli esperimenti di laboratorio prevedono la manipolazione di più di un fattore causale. Se i fattori manipolati in un esperimento sono più di uno, e il ricercatore non ritiene opportuno utilizzare una manipolazione intra-soggetti, il disegno che si utilizza per determinare le condizioni sperimentali è il cosiddetto *disegno fattoriale*. Il disegno fattoriale richiede la realizzazione di tante condizioni sperimentali quante sono le combinazioni possibili dei fattori causali che vengono manipolati nell'esperimento. Riprendendo l'esempio utilizzato in precedenza, supponiamo che il programma di intervento per prevenire i comportamenti a rischio preveda un secondo tipo di stimolo: l'esposizione a materiale audiovisivo realizzato con l'intento di promuovere l'impegno scolastico. I fattori causali manipolati saranno così due, la destinazione di una parte di ore a specifiche attività di gruppo e l'esposizione a materiale audiovisivo. I due fattori sono indipendenti, quindi ciascuno dei due può essere presente o assente in una data condizione. Di conseguenza le condizioni possibili in questo caso sono 2×2 : nessun intervento, solo attività di gruppo, solo materiale audiovisivo, oppure entrambi. Se per ipotesi uno dei due fattori potesse essere ulteriormente variato, prevedendo per esempio un numero di ore variabile, definito come alto o basso, da dedicare all'attività di gruppo, il numero di condizioni per il solo fattore attività di gruppo salirebbe a tre (assente, basso e alto) e il numero di condizioni richieste dal disegno fattoriale salirebbe a 3×2 . Naturalmente, all'aumentare del numero di condizioni sperimentali previste, aumenta il numero di confronti che andranno effettuati per valutare i risultati, e di conseguenza aumenta rapidamente il numero di partecipanti necessario per ottenere una potenza statistica sufficiente.

In qualche caso gli esperimenti possono avere un disegno misto, nel quale alcuni fattori sono sottoposti a manipolazione inter-soggetti ed altri a manipolazione intra-soggetti. Per esempio Simpson (2003) condusse una serie di esperimenti il cui obiettivo era di osservare se il tasso di cooperazione in alcuni specifici tipi di interazione fosse diverso fra uomini e donne, utilizzando una serie di «dilemmi sociali» appositamente disegnati per isolare singole motivazioni cognitive per il comportamento non cooperativo, quali il ti-

more della defezione altrui e l'avidità¹⁴. Simpson di fatto utilizzò un disegno misto, composto da un fattore causale inter-soggetti, il genere, e uno intra-soggetti, il tipo di dilemma sociale, che presentava tre condizioni possibili (cooperazione ostacolata dal timore, cooperazione ostacolata dall'avidità e cooperazione ostacolata da entrambe). I risultati dell'esperimento non evidenziarono differenze significative nelle scelte cooperative di uomini e donne quando la scelta di non cooperare era motivata dalla paura, mentre nelle interazioni in cui la scelta non cooperativa era motivata solo dall'avidità il tasso di cooperazione degli uomini risultò significativamente più basso di quello delle donne.

5.3 Gli esperimenti e la teoria. Il modello della scelta razionale e l'economia sperimentale

In un recente manuale dedicato alla costruzione di esperimenti di laboratorio nelle scienze sociali i sociologi Willer e Walker (2007) adottarono una posizione piuttosto radicale. Il loro manuale, infatti, è imperniato sulla distinzione draconiana fra *esperimenti a base empirica* ed *esperimenti a base teorica*. Le radici filosofiche del metodo sperimentale che abbiamo descritto nel primo paragrafo per Willer e Walker pertengono in modo specifico a quelli che loro definiscono esperimenti a base empirica. In linea con i canoni di Mill e successivamente con l'inferenza statistica formalizzata da Fischer, lo scopo degli esperimenti a base empirica è la scoperta di regolarità empiriche (che Mill credeva assolute e Fischer successivamente riformulò in termini probabilistici). Al contrario, gli esperimenti a base teorica, che gli autori illustrano attingendo copiosamente a esempi di classici esperimenti condotti nelle scienze fisiche e naturali, hanno lo scopo esplicito di testare ipotesi che sono derivate da modelli teorici formali. Quindi gli esperimenti a base teorica non producono scoperte, al contrario i risultati di questi esperimenti sono anticipati dalla teoria. La differenza fra i due tipi di esperimenti può essere evidenziata utilizzando i criteri guida per i due tipi di esperimenti proposti dai due autori. I criteri per la costruzione di un esperimento a base empirica sono i seguenti:

1. Creare due situazioni sperimentali che siano inizialmente il più possibile identiche
2. Introdurre una singola differenza fra le due situazioni e osservare il risultato
3. Utilizzare l'inferenza per attribuire il risultato all'effetto della specifica differenza introdotta

¹⁴ Il termine dilemma sociale si usa in sociologia e psicologia sociale per indicare una classe di interazioni costituita da situazioni in cui i soggetti, a coppie, devono decidere fra due opzioni di scelta alternative, descritte in modo astratto, i cui esiti, solitamente costituiti da piccoli compensi monetari, dipendono dalla combinazione delle scelte effettuate dai due attori coinvolti nell'interazione. I dilemmi sociali sono interazioni di questo tipo in cui gli individui sono messi di fronte a motivazioni alternative e confliggenti, che rendono le loro decisioni particolarmente problematiche e per questo scientificamente interessanti. Per una definizione più approfondita del termine si veda il Riquadro 5.3.

4. Inferire regolarità probabilistiche solo se è altamente improbabile che i risultati siano dovuti al caso

I criteri indicati dai due autori per la costruzione di un esperimento a base teorica invece sono i seguenti:

1. Derivare uno (o più) modelli dalla teoria che si intende testare
2. Usare la teoria per generare predizioni collegando le condizioni iniziali a quelle finali
3. Costruire delle repliche dei modelli, stabilire le condizioni iniziali e osservare le condizioni finali
4. Confrontare i risultati alle predizioni e decidere se la teoria è supportata

I criteri relativi alla costruzione degli esperimenti a base empirica sono interamente silenti in merito alla scelta del fattore che si intende manipolare, perché di fatto per gli autori questo tipo di esperimenti è basato su un approccio di tipo quasi esclusivamente induttivo (vedi Riquadro 1.7). Al contrario, gli esperimenti a base teorica hanno un carattere fortemente deduttivo. Quando, nel primo dei criteri sopra elencati, parlano di «modelli», gli autori si riferiscono a modelli astratti, spesso costituiti da sistemi di equazioni matematiche. Per loro di conseguenza, il disegno dell'esperimento consiste essenzialmente nella costruzione di una o più «repliche» concrete di questi modelli astratti all'interno del laboratorio. Infatti, benché anche i loro esperimenti prevedano generalmente la presenza di più condizioni sperimentali, non parlano di gruppo di controllo vs gruppo sperimentale, in riferimento agli esperimenti a base teorica. La caratteristica più importante dei modelli teorici che gli autori hanno in mente è la capacità di produrre predizioni *puntuali* di ciò che dovrebbe accadere durante l'esperimento, a condizione che la replica utilizzata corrisponda in modo sufficiente al modello sul quale è basata. Negli esperimenti a base teorica il confronto fra le predizioni del modello astratto e i risultati ottenuti con la replica sostituiscono il confronto fra i risultati ottenuti nel gruppo sperimentale e quelli ottenuti nel gruppo di controllo. Nonostante vi sia del merito nella distinzione proposta da Willer e Walker, è facile intuire che non sono molti i fenomeni studiati in psicologia per i quali siano disponibili modelli teorici in grado di fornire predizioni puntuali.

D'altra parte gli esperimenti a base empirica non hanno sempre un carattere eminentemente induttivo, anzi, nella maggior parte dei casi sono anch'essi basati su modelli teorici, ma le teorie da cui prendono origine sono costituite da insiemi di assunti e proposizioni, e non da sistemi di equazioni, si veda per esempio l'esperimento sulla diffusione del disordine descritto nel Riquadro 5.2. Tuttavia, dal momento che la maggior parte dei modelli teorici utilizzati in psicologia e in sociologia non fornisce predizioni puntuali, la relazione fra le ipotesi da essi derivate e il disegno sperimentale ha caratteristiche diverse da quelle relative agli esperimenti basati su modelli teorici formali. Negli esperimenti a base empirica le teorie guidano la scelta dei fattori da manipolare negli esperimenti e consentono la formulazione di ipotesi relative alla *direzione* (positiva o negativa) dell'effetto atteso. Infatti in molti di questi esperimenti le ipotesi sono verificate per mezzo di test statistici ad una coda (sulla relazione fra teoria e utilizzo di test statistici a una o due

code si veda la nota 9 di questo capitolo). Utilizzando ancora una volta lo stesso esempio del paragrafo precedente, se vogliamo valutare gli effetti di un programma di prevenzione dei comportamenti a rischio per studenti delle scuole secondarie, disporremo di una qualche teoria sulla base della quale testeremo l'ipotesi che l'incidenza dei comportamenti a rischio sia minore in conseguenza degli effetti del programma di prevenzione, mentre tralascieremo l'ipotesi contraria, cioè che l'effetto del programma di prevenzione sia un aumento dei comportamenti a rischio. L'ipotesi contraria potrebbe comunque verificarsi empiricamente, se il programma risultasse tanto fallimentare da generare un peggioramento, ma la nostra teoria ci consentirebbe comunque di usare un test statistico meno conservativo e quindi più potente, concentrandoci sull'ipotesi ritenuta più plausibile. La distinzione fra esperimenti a base empirica e teorica è comunque utile, e ci consentirà nella parte seguente del capitolo di analizzare un'ultima famiglia di esperimenti che hanno un legame particolarmente forte con modello teorico preciso, il modello della scelta razionale.

Gli esperimenti di laboratorio vengono solitamente associati alla psicologia sociale, ma nelle scienze sociali le comunità scientifiche principali all'interno delle quali la tecnica sperimentale è comune, sono almeno quattro, ed esistono importanti differenze di carattere metodologico fra le diverse comunità. La più nota è quella degli psicologi sociali in senso lato, ricercatori che hanno una formazione di tipo psicologico, sono collocati all'interno di dipartimenti di psicologia e sono membri delle associazioni professionali degli psicologi. Una seconda comunità, più piccola, è quella della «psicologia sociale sociologica», costituita dai sociologi che si occupano di processi di gruppo. Questi ultimi hanno perlopiù una formazione di tipo sociologico, sono collocati all'interno di dipartimenti di sociologia di università americane e membri di associazioni professionali dei sociologi, ma hanno spesso un'identità professionale mista. Sono gli eredi degli interazionisti simbolici ma sono fortemente influenzati dai classici della psicologia sociale sperimentale, fra i quali per esempio i lavori di Solomon Asch e Stanley Milgram¹⁵. In Europa la situazione è un po' più semplice, la psicologia sociale nella maggior parte dei paesi europei è contenuta all'interno dei dipartimenti di psicologia, però vi è una comunità scientifica di sociologi che utilizzano il metodo sperimentale, relativamente piccola ma piuttosto citata ed influente, sono i sociologi che aderiscono all'approccio della sociologia analitica (Hedström e Bearman 2009, Barbera 2004, Manzo 2014) e al modello teorico della scelta razionale. La caratteristica principale della sociologia analitica è che il tipo di spiegazione che propone è basato su teorie micro-fondate dell'azione (di cui la teoria della scelta razionale costituisce un prominente esempio) che danno conto dei fenomeni macro descrivendoli come «conseguenze inattese» o «effetti emergenti» delle interazioni degli attori coinvolti. Coerentemente con l'enfasi sulla componente micro della spiegazione, le ricerche che si concentrano su modelli di interazione e li studiano mediante gli esperimenti di laboratorio sono piuttosto comuni nell'ambito della sociologia analitica, e

¹⁵ L'interazionismo simbolico è una corrente sociologica sviluppatasi negli Stati Uniti verso la metà del secolo scorso, a partire dal lavoro di George Mead, che attribuiva un ruolo centrale alle interazioni sociali nella produzione di significati.

ancor di più fra i ricercatori che adottano l'approccio della scelta razionale. A causa della comunanza di interessi scientifici e della stretta relazione che intercorre fra il modello della scelta razionale e quello dell'homo oeconomicus (di cui parleremo più avanti), i sociologi della scelta razionale utilizzano disegni e procedure sperimentali mutuati da, o molto simili a, quelli utilizzati dalla quarta comunità scientifica che utilizza in modo preponderante il metodo sperimentale: quella dell'economia sperimentale.

Gli esperimenti economici e quelli dei sociologi della scelta razionale appartengono alla categoria degli esperimenti a base teorica, ma utilizzano una procedura che differisce in modo sostanziale da quella utilizzata dagli psicologi sociali, anche nei casi in cui l'oggetto di studio è il medesimo. Gli economisti e i sociologi della scelta razionale si sono occupati soprattutto di studiare i meccanismi decisionali in situazioni in cui gli esiti delle interazioni dipendono dalla combinazione delle decisioni individuali degli individui coinvolti, queste situazioni sono definite *interazioni strategiche*. In particolare, il problema che desta maggior interesse è quello della cooperazione in interazioni in cui il perseguimento dell'interesse individuale favorirebbe la non cooperazione. Lo studio delle interazioni strategiche si avvale di strumenti e modelli sviluppati in teoria dei giochi, il più famoso dei quali è il dilemma del prigioniero (si veda il Riquadro 5.3).

Gli strumenti di teoria dei giochi sono stati utilizzati anche dagli psicologi sociali che si occupano di cooperazione, altruismo, e in generale decisioni nei dilemmi sociali, si veda per esempio l'esperimento di Simpson (2003), descritto nel paragrafo precedente, in generale tuttavia gli psicologi sono più interessati alle motivazioni e ai meccanismi cognitivi, che non allo studio della razionalità. Il modello teorico più semplice utilizzato nella teoria dei giochi è quello dell'homo oeconomicus caratterizzato dall'assunto fondamentale che le decisioni individuali siano orientate unicamente alla massimizzazione dell'utilità attesa. L'utilità attesa può essere rappresentata con una funzione matematica e, nel modello dell'homo oeconomicus, si assume che essa dipenda unicamente dai profitti materiali che gli individui possono ottenere nelle interazioni. La teoria della scelta razionale, invece, comprende una classe di modelli che assumono che l'azione umana sia orientata allo scopo e che, nel perseguire i propri scopi, gli attori siano guidati dalla razionalità strumentale.

Come nel modello dell'homo oeconomicus, che rappresenta il capostipite dei modelli di scelta razionale, l'azione orientata allo scopo può essere rappresentata con una funzione di utilità attesa, ma a differenza del modello dell'homo oeconomicus, nei modelli di scelta razionale la funzione di utilità non dipende necessariamente solo dai profitti individuali, ma può includere anche motivazioni di tipo cognitivo o psicologico, o legate alla situazione contestuale¹⁶.

Gli esperimenti sociologici condotti nell'ambito della teoria della scelta razionale, e

¹⁶ Questa definizione della teoria della scelta razionale corrisponde a quella che Opp (2013a) chiama versione «ampia» (wide) di questa teoria, in contrapposizione ad una versione «ristretta» (narrow) che di fatto corrisponde al modello dell'homo oeconomicus. Il dibattito scientifico sulla relazione fra la teoria della scelta razionale, il modello dell'homo oeconomicus e la sociologia analitica è molto ampio e non privo di aspetti controversi (per esempio si vedano Opp 2013a, 2013b e Manzo 2013; una rassegna di alcuni di questi modelli teorici, applicati a problemi di cooperazione, si può trovare invece in Barrera 2008, 2014).

Riquadro 5.3

La teoria dei giochi e il dilemma del prigioniero

LA TEORIA DEI GIOCHI

Il termine «giochi» indica una tipologia di interazioni strategiche, ogni gioco rappresenta un modello stilizzato di un tipo di interazione con caratteristiche diverse; la «teoria» invece è costituita dall'insieme di regole e assunti che vengono utilizzati per «risolvere» il gioco individuandone la «soluzione» (Camerer 2003). La Soluzione di un gioco corrisponde ad una ipotesi relativa a come gli individui dovrebbero comportarsi se gli assunti della teoria sono corretti e se l'interazione in cui sono coinvolti riflette le caratteristiche del gioco utilizzato per analizzarla. La teoria dei giochi ha moltissime applicazioni nei campi scientifici più disparati (in matematica, economia e finanza, sociologia, biologia, psicologia, scienza della politica, informatica ed intelligenza artificiale).

IL DILEMMA DEL PRIGIONIERO

Il dilemma del prigioniero è il più famoso e studiato fra i modelli di interazione strategica e prende il nome dalla storiella con cui fu reso popolare negli anni 50 dal matematico canadese Albert Tucker. Due criminali, accusati di aver commesso un reato, vengono arrestati dalla polizia e posti in due celle separate dove, durante l'interrogatorio, viene offerta loro la possibilità di confessare, fornendo alla polizia le prove per incriminare il loro complice, in cambio della sospensione delle accuse nei loro confronti. Entrambi si trovano così di fronte al dilemma di tradire il proprio complice per salvare se stessi oppure tacere ed essere incriminati. Le due alternative hanno le seguenti possibili conseguenze:

- Se uno solo dei due confessa, evita la pena, ma il complice viene condannato a 7 anni
- Se entrambi confessano, vengono condannati entrambi a 6 anni ciascuno
- Se nessuno dei due confessa, vengono condannati solo per un reato minore con una sentenza di 1 anno ciascuno

Assumendo che ai due criminali interessi soltanto minimizzare la propria condanna, il gioco ha una soluzione molto semplice: la scelta di tradire il complice e confessare garantisce una sentenza più leggera, sia nel caso in cui l'altro confessi, sia nel caso in cui taccia. In teoria dei giochi, una opzione con queste caratteristiche è detta *strategia dominante*. Un secondo concetto utilizzato per individuare la soluzione di un gioco è l'equilibrio di Nash (così chiamato in onore del suo scopritore, John Nash). L'equilibrio di Nash è quell'esito dell'interazione in cui a nessuno dei partecipanti conviene cambiare *unilateralmente* l'opzione scelta. Nel dilemma del prigioniero l'esito in cui entrambi tacciono è un'equilibrio di Nash. Non tutti i giochi possiedono una strategia dominante, ma tutti i giochi in cui il numero delle situazioni possibili è finito possiedono almeno un equilibrio di Nash (talvolta più di uno). Il dilemma del prigioniero è uno dei modelli più usati per lo studio del problema della cooperazione, perché rappresenta una situazione in cui la scelta cooperativa (tacere) è contraria agli interessi individuali, la cui massimizzazione porterebbe alla scelta opposta, quella di confessare. In sociologia e psicologia sociale le interazioni di questo tipo sono definite *dilemmi sociali*.

anche quelli degli economisti, rientrano fra gli esperimenti a base teorica, infatti le ipotesi testate sono formalmente derivate da modelli teorici di cui le interazioni studiate in laboratorio rappresentano una replica. Tuttavia, i modelli di scelta razionale, e in particolare il principale fra questi, il modello dell'*homo oeconomicus*, vengono spesso utilizzati dai ricercatori soprattutto per la loro utilità come modelli *normativi*, e non per la loro accuratezza come modelli *descrittivi*. Il modello della scelta razionale è l'equivalente sociologico del modello del gas perfetto in fisica: non è particolarmente utile per descrivere il comportamento degli individui nelle interazioni sociali, dal momento che le situazioni in cui il comportamento degli individui riflette quello previsto dal modello sono poco comuni, ma è molto utile come modello di riferimento con il quale confrontare il comportamento effettivo dei soggetti nelle interazioni studiate, in particolare per investigare le motivazioni cognitive dei comportamenti non conformi al modello teorico.

Dal punto di vista metodologico, lo sforzo principale dei ricercatori è stato quello di costruire in laboratorio delle repliche (nella stessa accezione utilizzata da Willer e Walker) che rispecchiassero nel modo più fedele possibile i modelli di interazione astratti della teoria dei giochi. Alcuni fra i giochi maggiormente studiati sono descritti nel Riquadro 5.4. Nello sforzo volto a rendere le repliche identiche ai modelli teorici su cui gli esperimenti si basano, gli economisti hanno prodotto una serie di innovazioni o modifiche delle procedure sviluppate e consolidate in psicologia sociale che descriveremo di seguito, che sono state generalmente adottate anche dai sociologi della scelta razionale. Alcune differenze minori, per esempio una minor riluttanza ad usare la manipolazione intra-soggetti da parte degli economisti, sono già state menzionate nel paragrafo precedente. Una elaborata analisi dei punti di maggior disaccordo fra le due scuole, condotta dal punto di vista degli economisti, si può trovare in un lungo articolo pubblicato su *Brain & Behavioral Sciences* da Hertwig e Ortmann (2001), corredato da una serie di open commentaries da parte di ricercatori di tutte le discipline sociali sperimentali.

Definizione delle condizioni sperimentali. Come osservato da Willer e Walker (2007), gli esperimenti a base teorica non necessitano di una vera e propria condizione di controllo perché utilizzano le predizioni puntuali dei modelli teorici sui quali sono basati, come termine di paragone per i risultati dell'esperimento.

Ciononostante, anche gli esperimenti a base teorica, compresi quelli di economia sperimentale, di norma presentano almeno due condizioni. La differenza, rispetto agli esperimenti a base empirica, è che ciò che distingue le due condizioni non è semplicemente la presenza/assenza del fattore manipolato, ma una modifica di quelle che Willer e Walker chiamano «condizioni iniziali» oppure l'introduzione di un elemento innovativo non previsto nel modello originale su cui l'esperimento è basato. Per questo motivo, negli esperimenti di economia e scelta razionale, la condizione di controllo prende il nome di condizione di *riferimento* (baseline) ed è solitamente costituita dalla versione standard di uno dei giochi maggiormente noti e studiati, come quelli descritti nel Riquadro 5.4. Per esempio, nell'esperimento di Fehr e Gächter (2002) la condizione di riferimento è un classico gioco del bene pubblico, al quale, nella condizione sperimentale, viene aggiunta una modifica, la possibilità di utilizzare uno strumento sanzionatorio.

Incentivi monetari. L'uso degli incentivi monetari è uno dei capisaldi dell'economia

sperimentale (Hartwig e Ortmann 2001). La prassi di retribuire i soggetti partecipanti agli esperimenti è comune anche in psicologia, ma in psicologia la retribuzione non è necessariamente una funzione diretta delle decisioni prese dai partecipanti durante l'esperimento. Al contrario gli economisti credono fermamente che, affinché gli assunti del modello siano correttamente riprodotti nella replica, le decisioni devono sempre avere conseguenze reali. Queste conseguenze reali sono operativizzate per mezzo di incentivi monetari semplicemente perché il denaro ha la utile proprietà di essere un oggetto ragionevolmente apprezzato da tutti, e consente quindi di definire gli esiti di una interazione in modo tale che riflettano specularmente l'ordine di preferenza individuale postulato dal modello. Gli economisti sostengono che un esperimento privo di incentivi monetari non può fornire alcuna informazione utile in merito al modo in cui gli individui prendono le decisioni. Tuttavia, la gamma di situazioni e interazioni studiate dagli psicologi comprende anche molte situazioni in cui le azioni dei soggetti non sono trasformabili in decisioni economiche e per le quali l'applicazione di incentivi economici non è possibile. Generalmente, gli incentivi applicati dagli economisti sono peraltro abbastanza modesti (il compenso individuale per un esperimento della durata di circa un'ora è solitamente uguale a due/tre ore di lavoro pagate al salario orario minimo del paese in cui ha luogo l'esperimento), anche se occasionalmente sono stati condotti esperimenti con incentivi più elevati. I più noti fra questi sono certamente gli studi cross-culturali su norme, altruismo e cooperazione condotti da un team di antropologi ed economisti, guidato da Joseph Henrich, in una ventina di comunità locali – eterogenee rispetto all'organizzazione economica, all'ambiente geografico e al tipo di residenza – in vari paesi, la maggior parte dei quali in via di sviluppo. In questi esperimenti, basati su alcuni dei giochi descritti nel Riquadro 5.4, Henrich e colleghi utilizzarono un compenso pari al salario giornaliero nell'economia locale e osservarono una elevata eterogeneità nei comportamenti dei vari campioni (per una rassegna dei risultati più interessanti si veda Esminger e Henrich 2014)¹⁷.

Anonimità. Come detto in precedenza l'anonimità in doppio cieco, viene utilizzata dai ricercatori per eliminare il bias dello sperimentatore, e per evitare che i partecipanti si formino delle impressioni relative alle aspettative del ricercatore o degli altri partecipanti e il loro comportamento ne venga influenzato. Come per gli incentivi monetari, gli economisti utilizzano l'anonimità soprattutto allo scopo di rendere la replica dell'interazione identica a quella postulata dal modello. Uno degli assunti cruciali del modello dell'*homo oeconomicus*, e della teoria della scelta razionale nella sua forma più radicale, è che l'utilità degli attori sia unicamente una funzione dei propri esiti attesi, senza alcuna conside-

¹⁷ In un altro lavoro, ispirato dai loro esperimenti cross-culturali, Henrich e colleghi presentarono una copiosa rassegna dei risultati sperimentali riguardanti psicologia e comportamento umano, accumulati nelle varie discipline e basati in larghissima parte su campioni di soggetti provenienti da società che gli autori definiscono, con un efficace acronimo, WEIRD (western, educated, industrialized, rich, democratic), cioè occidentali, istruite, industrializzate, ricche e democratiche (weird = strano, molto insolito). In questo articolo gli autori conclusero che i membri delle società WEIRD sono fra le popolazioni meno rappresentative che si possano utilizzare per trarre conclusioni riguardanti gli esseri umani in generale (Henrich, Heine e Norenzayan 2010).

Riquadro 5.4**Alcuni fra i giochi più studiati****IL GIOCO DELLA FIDUCIA**

Il gioco della fiducia rappresenta un problema di cooperazione sequenziale, anziché simultaneo. Il giocatore che sceglie per primo, detto “trustor” sceglie fra l’alternativa di riporre la propria fiducia nel secondo giocatore, detto “trustee”, oppure di non farlo. Se il trustor decide di non fidarsi, il gioco termina, se invece si fida, il trustee può decidere se onorare la fiducia ricevuta oppure abusarne. Gli esiti possibili sono 3 (la fiducia non viene posta, la fiducia viene posta e abusata, la fiducia viene posta e onorata) e sono stabiliti in modo da determinare le seguenti preferenze: il trustor preferisce che la fiducia sia onorata piuttosto che abusata, ma preferisce non fidarsi piuttosto che la propria fiducia sia abusata; il trustee preferisce abusare piuttosto che onorare la fiducia, ma preferisce onorare la fiducia piuttosto che la fiducia non venga posta. In equilibrio la soluzione del gioco è che il trustor non pone la fiducia, perché altrimenti verrebbe abusata.

IL GIOCO DEL BENE PUBBLICO

Il gioco del bene pubblico rappresenta un problema di cooperazione che coinvolge più di due giocatori. In questo gioco i partecipanti dispongono di una dotazione iniziale di risorse e possono decidere di investirla in un bene pubblico. Il bene pubblico, se almeno uno dei partecipanti contribuisce, produce un surplus che viene diviso in parti uguali fra i partecipanti. Dal punto di vista collettivo è più conveniente che tutti i partecipanti contribuiscano con l’intera dotazione alla produzione del bene pubblico, ma dal punto di vista individuale è più conveniente tenersi la dotazione e approfittare della contribuzione altrui. In equilibrio la soluzione di questo gioco è che nessuno contribuisce e il bene pubblico non viene prodotto. Fehr e Gächter (2002) utilizzarono questo gioco nell’esperimento descritto nel secondo paragrafo, aggiungendo, al termine dell’interazione, la possibilità per gli attori di punire uno degli altri giocatori. La punizione comporta un costo per chi la effettua e ha l’effetto di sottrarre una parte delle risorse possedute a chi la riceve.

IL GIOCO DEL DITTATORE

Il gioco del dittatore è uno strumento molto utilizzato per studiare il comportamento altruista. In questo gioco un giocatore (il dittatore) dispone di una dotazione che può decidere di spartire con il secondo giocatore, dividendola a suo piacimento, oppure può decidere di tenere tutta la dotazione per sé. Il secondo giocatore si limita a ricevere quello che il dittatore decide di offrirgli. In equilibrio il dittatore dovrebbe decidere di tenere tutta la dotazione per sé.

IL GIOCO DELL’ULTIMATUM

Il gioco dell’ultimatum è simile al gioco del dittatore con una modifica: il secondo giocatore, detto ricevente, può decidere di respingere la spartizione proposta dal primo giocatore, detto proponente, e in questo caso entrambi restano a mani vuote. Se invece accetta, la dotazione viene divisa fra i due, come stabilito dal proponente. In equilibrio il proponente dovrebbe offrire al ricevente la porzione minima possibile della dotazione, e il ricevente dovrebbe accettare la proposta.

IL GIOCO DELLA PUNIZIONE DA PARTE TERZA

Anche questo è simile al gioco del dittatore con una differenza: vi è un terzo giocatore che assiste alla decisione del dittatore e può decidere di punirlo. La punizione comporta un costo per il terzo giocatore e ha l’effetto di ridurre la dotazione posseduta dal dittatore. In equilibrio la punizione non dovrebbe mai essere usata e il dittatore dovrebbe tenere per sé tutta la dotazione.

razione per gli esiti ottenuti dagli altri partecipanti all'interazione. L'anonimità nelle interazioni, come dimostrato anche da alcuni degli esperimenti di Milgram sull'obbedienza all'autorità (si veda il Riquadro 1.3), ha sicuramente l'effetto di rendere gli individui più indifferenti alle conseguenze delle proprie azioni subite da altri. Quindi, negli esperimenti economici l'anonimità ha principalmente lo scopo di evitare che le decisioni dei partecipanti siano influenzate da sentimenti di simpatia o sensi di colpa nei confronti degli altri partecipanti (manipolazione in cieco) o nei confronti del ricercatore (manipolazione in doppio cieco).

Realismo. Il realismo rappresenta, invece, uno dei capisaldi per gli psicologi sociali (Willer e Walker 2007). Poiché il campo di ricerca degli psicologi sociali riguarda in particolare elementi psicologici, motivazionali e cognitivi del comportamento umano nelle interazioni sociali, la condizione di artificialità della situazione, che è tipica negli esperimenti di laboratorio, deve essere ridotta ad ogni costo, anche perché alcune delle risposte comportamentali più interessanti si manifestano in modo involontario o inconsapevole, quando il soggetto è convinto che la situazione in cui si trova sia perfettamente reale. Al contrario, l'obiettivo degli economisti è quello di creare delle condizioni in cui i soggetti possano operare come dei computer, senza subire alcuna pressione o influenza ambientale che susciti in loro emozioni o impulsi che potrebbero alterarne il comportamento. Per questo motivo gli esperimenti di economia sono basati su un copione standardizzata caratterizzata dal massimo grado di astrattezza e priva di qualunque riferimento verbale o semantico che possa sollecitare l'attivazione di meccanismi diversi dalla pura razionalità. Hertwig e Ortmann (2001) sostengono che il mancato uso, da parte degli psicologi, di un copione standardizzata di questo tipo potrebbe indurre i soggetti a «improvvisare» assumendo comportamenti erratici e casuali. Probabilmente l'idea che l'uso di un copione standardizzata e astratta sia sufficiente a impedire completamente l'elaborazione di interpretazioni cognitive della situazione da parte dei soggetti è utopistica, nondimeno questo tipo di procedura è universalmente adottato negli esperimenti di economia.

Ripetizione. Nella loro rassegna dei punti di disaccordo metodologico fra psicologi ed economisti Hertwig e Ortmann (2001) inclusero anche l'uso delle prove ripetute seguite da feedback. In alcuni esperimenti di economia è stato osservato che l'utilizzo della ripetizione seguita da feedback produce risultati che nel tempo tendono a aumentare il grado di corrispondenza fra i risultati degli esperimenti e le predizioni dei modelli. Questo aspetto è generalmente poco rilevante per gli psicologi per due motivi. Primo, poiché gli esperimenti di psicologia non si basano su modelli formali che forniscono predizioni puntuali, la ripetizione di un compito sperimentale ha senso soltanto se il compito presenta un grado di difficoltà tale per cui è presumibile che i risultati più interessanti si possano ottenere solo dopo che i soggetti hanno effettuato alcune prove, ed infatti in questo tipo di esperimenti l'uso di prove ripetute non è infrequente nemmeno fra gli psicologi sociali, per esempio si vedano gli esperimenti sulla teoria dello scambio sociale (Willer 1999). Secondo, nella maggior parte degli esperimenti condotti in psicologia, la prima risposta, non mediata da alcun processo di apprendimento o familiarizzazione con la situazione sperimentale, è la più interessante (si pensi agli esperimenti di Asch, descritto nel Riqua-

dro 5.1 di questo capitolo, Milgram o Zimbardo, descritti nel Riquadro 1.3); eventuali ripetizioni, anche qualora fornissero esiti diversi, non sarebbero più ugualmente spontanee.

Uso dell'inganno. La questione metodologica più controversa e quella su cui è stato scritto di più è sicuramente l'uso dell'inganno, definito come l'utilizzo intenzionale ed esplicito di informazioni false o ingannevoli relative ai fatti o alle persone coinvolte nell'esperimento (McDaniel and Starmer 1998). Negli esperimenti di psicologia sociale l'inganno viene usato frequentemente allo scopo di aumentare il grado di realismo, che a sua volta permette di migliorare la validità degli esperimenti. Al contrario nell'economia sperimentale l'uso dell'inganno costituisce un vero e proprio tabù¹⁸. Sebbene l'uso dell'inganno abbia innegabilmente anche dei risvolti etici, la controversia riguarda aspetti puramente pragmatici della ricerca sperimentale, infatti gli economisti sostengono che l'inganno induca i partecipanti a diffidare dei ricercatori ed assumere un atteggiamento meno cooperativo, minacciando quindi la validità dei risultati sperimentali (McDaniel e Starmer 1998, Hertwig e Ortmann 2001; in merito all'importanza della cooperazione fra osservatore e soggetti osservati, si veda anche il capitolo 1 di questo volume). L'utilizzo di convenzioni confliggenti in merito all'uso dell'inganno nelle diverse discipline ha conseguenze problematiche per la produzione scientifica perché, data la natura interdisciplinare della ricerca su temi come la cooperazione, l'altruismo e in generale le interazioni strategiche, articoli e progetti di ricerca che utilizzano l'inganno rischiano di venire rigettati da referee che lo ritengono inammissibile. A dispetto del fatto che la questione venga posta in termini pragmatici e non etici, il dibattito talvolta assume toni ideologici. Tuttavia, se la controversia ha effettivamente una natura pragmatica legata a possibili conseguenze dell'inganno sull'atteggiamento cooperativo dei partecipanti, la questione potrebbe essere risolta empiricamente. Barrera e Simpson (2012) realizzarono un esperimento che aveva esattamente questo scopo.

L'esperimento consisteva di due parti, nella prima parte gli studenti, a coppie, parteciparono ad una interazione modellata sul dilemma del prigioniero in cieco. I ricercatori divisero i 153 partecipanti in due gruppi mediante assegnazione causale. Nel gruppo di controllo i soggetti effettivamente interagivano con un altro partecipante, nel gruppo sperimentale l'interazione era simulata: mentre i soggetti credevano di interagire con un'altra persona, la decisione del partner veniva determinata mediante il lancio di un dado. Al termine dell'esperimento i partecipanti vennero pagati e quelli del gruppo sperimentale vennero informati dell'inganno subito. Tre settimane dopo i ricercatori richiamarono i partecipanti per la seconda parte dell'esperimento, che fu presentata ai soggetti come un nuovo esperimento. Nella seconda parte i ricercatori rilevarono le variabili dipendenti dello studio, utilizzando altri modelli di interazione, fra cui il gioco della fiducia e il gioco del dittatore (si veda il Riquadro 5.4), e un breve questionario. Siccome notoriamente la

¹⁸ Hertwig e Ortmann riportarono che, nel 2001, circa un terzo degli articoli pubblicati sulla rivista più prestigiosa di psicologia sociale, il *Journal of Personality and Social Psychology*, presentavano esperimenti realizzati utilizzando qualche forma di inganno, mentre gli editori delle più importanti riviste di economia non accettano articoli che utilizzino l'inganno e in tutta la letteratura di economia sperimentale gli esperimenti condotti utilizzando l'inganno possono essere contati sulle dita di due mani (Hertwig e Ortmann 2001).

cooperazione e l'altruismo tendono a ridursi se i soggetti credono che l'interazione sia fittizia e il loro partner sia in realtà simulato, gli effetti negativi dell'inganno, postulati dagli economisti, avrebbero dovuto manifestarsi sotto forma di offerte più basse nel gioco del dittatore e minor grado di fiducia nel gioco della fiducia, da parte dei soggetti che erano stati ingannati nella prima parte dell'esperimento. I risultati mostrarono che, sebbene i soggetti ingannati credevano che l'inganno fosse usato più spesso dei soggetti non ingannati, il comportamento dei partecipanti dei gruppi non differiva in modo significativo rispetto a nessuna delle misure utilizzate.

Mentre è auspicabile che le controversie di natura metodologica fra ricercatori di campi diversi che studiano gli stessi fenomeni si attenuino, adottando un approccio pragmatico che consenta di aumentare le collaborazioni interdisciplinari, nel rispetto delle rispettive posizioni, esiste un punto sul quale tutti i ricercatori che utilizzano il metodo sperimentale si trovano d'accordo. Il metodo sperimentale facilita l'accumulazione incrementale della conoscenza mediante la *replicazione*. Una parte considerevole degli esperimenti condotti in tutte le discipline viene infatti realizzata a partire da procedure consolidate, replicando disegni sperimentali già testati da altri, sui quali di volta in volta vengono inserite modifiche che generano nuove condizioni sperimentali.

Domande di autovalutazione

- 5.1 Che relazione c'è fra i canoni di Mill e il metodo sperimentale?
- 5.2 Qual è la differenza fra errore del primo e del secondo tipo e quali sono le implicazioni dei due tipi di errore per il processo di inferenza causale?
- 5.3 Quali sono le principali fonti di disturbo della validità interna di un esperimento e quali caratteristiche del disegno sperimentale che permettono di eliminare o ridurre l'effetto di queste fonti di disturbo?
- 5.4 Qual è la differenza fra un disegno sperimentale inter-soggetti e intra-soggetti e come differiscono le condizioni dell'esperimento nei due casi?
- 5.5 Cosa si intende per disegno fattoriale?
- 5.6 Come differiscono gli esperimenti cosiddetti «a base empirica» da quelli «a base teorica» e come vengono definite le condizioni dell'esperimento nei due tipi?
- 5.7 Quali sono le principali differenze metodologiche introdotte dall'economia sperimentale?